Amendments to the Claims

Please amend claim 1 as follows:

Listing of the claims

1. (Currently Amended) A computer readable medium storing a computer program to perform method steps for execution by a processor, the method steps comprising:

generating a synthetic waveform for each of N textual transcriptions of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition system and represent N-best textual transcription hypotheses of the original waveform;

for each synthetic waveform,

time-aligning feature vectors of the synthetic waveform with feature vectors of the original waveform at a phoneme level;

computing a mean of the feature vectors which align to each phoneme for the original waveform and the synthetic waveform;

computing a distance measure between each phoneme mean of the original waveform and the synthetic waveform;

summing the distance measures to generate an overall distance measure representing a distance between the original waveform and the synthetic waveform;

comparing scores based on the overall distance measure between the synthetic waveform and the original waveform, an acoustic model score of a corresponding textual transcription of the synthetic waveform, and a language model score of the corresponding textual transcription to determine a corresponding one of the N-best textual transcriptions; and

selecting for output the determined N-best textual transcription ~~corresponding to the synthetic waveform having a smallest overall distance measure,~~

~~wherein generating the synthetic waveform includes adjusting the pitch of the synthetic waveform to flat using a pitch synchronous overlap and add (PSOLA) technique~~.

2-3. (Cancelled)

4. (Previously Presented) The computer readable medium of claim 1, wherein the alignment is performed using a Viterbi alignment process.

5-8. (Cancelled)

9. (Previously Presented) A method for recognizing speech, the method comprising the steps of:

generating a synthetic waveform for each of N textual transcriptions of an original waveform, wherein N is greater than 1 and the N textual transcriptions are generated by a speech recognition system and represent N-best textual transcription hypotheses of the original waveform;

for each synthetic waveform,

computing a distance measure between the synthetic waveform and the original waveform;

summing the distance measures to generate an overall distance measure representing a distance between the original waveform and the synthetic waveform;

generating a score S from the overall distance measure D, an acoustic model score A of the corresponding textual transcription for the synthetic wave, and a language model score L of the corresponding textual transcription, wherein the score S = -D + (a * A) + (b * L), and 'a' and 'b' are constants;

selecting for output one of the textual transcriptions corresponding to the synthetic waveform having the score that indicates the synthetic wave is closest to the original waveform.

10. (Previously Presented) The method of claim 9, further comprising:

aligning frames of the original waveform and frames of each synthetic waveform to a corresponding one of the N textual transcriptions; and

calculating the distance measure between the original waveform and each of the synthetic waveforms based on the corresponding alignments.

11. (Previously Presented) The method of claim 10, further comprising:

retrieving feature vectors corresponding to the original waveform; and

generating feature vectors for each synthetic waveform such that the feature vectors for the synthetic waveforms are-similar in structure to the feature vectors of the original waveform,

wherein the alignment is performed by time-aligning the feature vectors of the original waveform and the feature vectors of each synthetic waveform with the corresponding one of the N textual transcriptions.

12-13. (Cancelled)

14. (Previously Presented) The method of claim 9, further comprising:

computing a mean feature vector of all feature vectors comprising each aligned frame for both the original and Nth synthetic waveform, wherein the distance measure for each aligned frame is calculated by determining a distance between each means of the corresponding aligned frames.

15. (Previously Presented) An automatic speech recognition system, comprising:

a decoder for decoding an original waveform of acoustic utterances to produce N textual transcriptions, the N textual transcriptions representing N-best textual transcription hypotheses of the decoded original waveform;

a text to speech system generating a synthetic waveform for each of the N textual transcriptions;

a means to perform a speaker normalization on the original waveform to match vocal-tract characteristics of a speaker from whose data the TTS is derived; and

a comparator for comparing scores based on an overall distance measure between each synthetic waveform and the normalized original waveform, an acoustic model score of a corresponding textual transcription of the synthetic waveform, and a language model

score of the corresponding textual transcription to determine a corresponding one of the N-best textual transcriptions to output,

wherein the overall distance measures are computed by a processor:

computing a distance measure between the synthetic waveform and the normalized original waveform; and

summing the distance measures to generate an overall distance measure representing a distance between the normalized original waveform and the synthetic waveform, and

wherein N is greater than 1.

16. (Previously Presented) The system of claim 15, further comprising a feature analysis processor adapted to generate a set of feature vectors for the normalized original waveform and generate a set of feature vectors for each of the N synthetic waveforms.

17-19. (Cancelled)

20. (Previously Presented) The system of claim 15, further comprises:

means for aligning frames of the normalized original waveform and frames of each synthetic waveform to a corresponding one of the N textual transcriptions; and

means for calculating the distance measure between the normalized original waveform and each of the synthetic waveforms based on the corresponding alignments.

21. (Original) The system of claim 20, wherein the frames are aligned on a phoneme level.

22. (Previously Presented) The system of claim 20, wherein the means for calculating the distance measures comprises:

means for calculating an individual distance between each aligned frame of the original normalized waveform and each of the N synthetic waveforms; and

means for summing the individual distances of the aligned frames of the original normalized waveform and each synthetic waveform to compute the overall distance measures between the original normalized waveform and each synthetic waveform.